

Randomized Control Trial Evaluation of the Implementation of the PSA-DMF in Dane County, WI

Summary of Final Report, Supplemental

Respectfully Submitted

D. James Greiner

S. William Green Professor of Public Law, Harvard Law School
Faculty Director, Access to Justice Lab
Ph. D., Statistics, Harvard University
J.D., University of Michigan Law School

Renee L. Danser, Associate Director of Research and Strategic Partnerships
Ryan Halen, Data Analyst
Access to Justice Lab
Harvard Law School

Christopher L. Griffin, Jr., Director of Empirical & Policy Research
James E. Rogers College of Law
University of Arizona

Matthew Stubenberg, Senior Product Manager
Morrison Foerster

July 23, 2026

One-Page Discussion

This Summary supplements a January 25, 2024, overview of the Final Report of the Access to Justice Lab at Harvard Law School (A2J) Lab with respect to a randomized study investigating the use of the PSA-DMF System in Dane County, Wisconsin.

The A2J Lab is a research entity within Harvard Law School that implements randomized controlled trials (RCTs), the scientific gold standard in discovering what works, to study how to improve the legal system for individuals who cannot afford to hire lawyers. The PSA-DMF System is a quantitative instrument that uses age and criminal history information to, purportedly, classify arrested individuals according to their risk if released of failing to appear or of being rearrested/cited for a new crime, especially a violent new crime. In jurisdictions that have implemented the PSA-DMF, a judicial officer (in Dane County, a “Commissioner”) receives a one-page printout as an informational tool to aid in making the first decision for an in-custody arrestee regarding bail, release conditions, and predisposition monitoring. Arnold Ventures (AV), a philanthropic foundation, funded the creation of the PSA-DMF.

In the 2010s, the Dane County Community Justice Council (CJC) led the County in a multi-year effort to modernize and improve the fairness of the County’s criminal justice system, an effort that included amending Local Rule 206 to reduce the time from jail booking to in-custody first appearance. In 2017, the County used a starter grant from AV to implement the PSA-DMF and, simultaneously, pursued a partnership with the A2J Lab to conduct an RCT to assess whether the PSA-DMF reduced some or all of failure to appear (FTA), new criminal activity (NCA), new violent criminal activity (NVCA), measures of racial disparity, and predisposition incarceration. From mid-2017 to the end of 2019, the A2J Lab and the County randomized whether the Commissioner and the parties received the PSA-DMF printout at in-custody first appearance hearings. After a two-year follow-up period, the County delivered data from multiple information systems. The A2J Lab spent years analyzing the data and, with help, developing new statistical methodology.

In early 2024, the A2J Lab reported to the CJC that providing the Commissioner with the PSA-DMF printout at in-custody first appearance hearings modestly altered in-custody first appearance decisions in the direction of PSA-DMF recommendations but neither improved nor worsened criminal justice outcomes. The A2J Lab provided hypotheses as to why the PSA-DMF appeared to have no effect but advised waiting until the Lab completed counterpart RCTs in two other jurisdictions to stress test those hypotheses.

The A2J Lab completed the counterpart RCTs. They produced results functionally identical to those from Dane County. These results, along with statistical advances constructed during the past five years, made the A2J Lab firm in the following conclusions.

- 1) The PSA-DMF System neither improves nor worsens FTA, NCA, NVCA, racial fairness, or predisposition incarceration in the County. If the County still expends resources to produce the PSA-DMF, it may wish to reconsider redirecting those resources.
- 2) Two reasons most likely explain these findings.
 - a. For all but a very small fraction of individuals, imposing cash bail does not prevent FTA, NCA, or NVCA.
 - b. The PSA-DMF does not classify arrestees well according to risk of FTA, NCA, or NVCA.

Supplemental Summary Report

I. Introduction

This Summary supplements a January 25, 2024, overview of the Final Report of the A2J Lab with respect to a randomized study investigating the use of the PSA-DMF System in Dane County, Wisconsin. As of early 2024, the A2J Lab's fundamental findings were as follows. First, the availability of the PSA-DMF caused modest changes in Commissioner decisions at in-custody first appearance hearings in the direction of PSA-DMF recommendations, but there were no relevant changes (positive or negative) in FTA, NCA, NVCA, number of predisposition incarceration days, or measures of racial fairness. Thus, the PSA-DMF neither achieved hoped-for improvements in criminal justice outcomes nor inflicted the harms that some had feared. Second, the most likely reasons for this lack of change were (i) the PSA's inability to classify risk well, and (ii) the fact that bail itself did not prevent misconduct. In January of 2024, neither explanation was certain, but both represented the A2J Lab's best guesses.

In the past 30 months, the A2J Lab has completed and analyzed counterpart RCTs of the PSA-DMF in two other sites, Polk County (IA) and four¹ counties in Utah. The A2J Lab and a team led by statistical methodologist Kosuke Imai have developed six original research papers articulating new statistical techniques based on the data from these studies, some of which incorporated the power of large language models. The A2J Lab and Imai teams have applied these new techniques to the Dane, Polk, and Utah data.

The Polk and Utah sites produced substantively the same results as those from Dane. The statistical techniques shed new light on why the PSA-DMF had no effect, good or bad. As a result, the A2J Lab's level of certainty in its findings has increased. The key takeaways are as follows.

First, the PSA-DMF induced moderate changes in Dane (and Utah), but large changes in Polk, in in-custody first decisions with respect to bail. Those changes were in the direction of PSA-DMF recommendations.

Second, in all three sites, the PSA-DMF had no relevant effect, positive or negative, on FTA, NCA, NVCA, number of days of predisposition incarceration, or measures of racial fairness.

Third, for all but a small fraction of arrested individuals (well below five percent on average), imposing cash bail had no effect on the likelihood of FTA, NCA, or NVCA.

Fourth, the PSA-DMF did not classify arrested individuals well in terms of whether imposing cash bail would decrease risk of FTA, NCA, or NVCA.

Fifth, the lack of an effect on predisposition incarceration days likely stemmed from several factors. Increases in signature/own-recognizance-release (ROR) decisions (and assignments of low cash bail) that the PSA-DMF induced when risk scores were low were largely negated by corresponding decreases in signature/ROR decisions when risk scores were high. Moreover, released arrestees returned to jail (from NCA or rearrest after FTA) at rates high enough to blunt the benefit of any net increase in signature/ROR decisions.

The A2J Lab is now comfortable in suggesting that, if Dane County continues to expend resources in producing the PSA-DMF System printouts for in-custody first appearance hearings, it should consider whether such continued expenditure is justified.

II. Background

In the early 2010s, the CJCⁱⁱ led Dane County's efforts to adopt reforms to the County's system for administering individuals between arrest and disposition of criminal charges. One impetus for this effort was concern over perceived racial disparities in the criminal justice system, including during the predisposition period. Another was a desire to reduce, safely, reliance on incarceration. A third was to incorporate more evidence-based thinking into the County's predisposition system.

After compiling a report articulating options to pursue these and other goals,ⁱⁱⁱ the CJC implemented a suite of reforms. For example, in July 2016, Dane County courts amended local Rule 206, covering bail hearing procedure, to reduce the amount of time between booking and in-custody first appearance hearings.^{iv} Law enforcement also increased the use of citations as opposed to custodial arrest for some suspected misdemeanors and low-level felonies, resulting in fewer in-custody first appearance hearings overall.^v In 2017, the Courts reduced the use of drug and alcohol testing and GPS monitoring as a condition of release.^{vi} In 2018, the County implemented state policy regarding the automatic removal of certain criminal records from public-facing websites, a process known as "sunsetting,"^{vii} as well as systems of voluntary text message reminders for court appearance dates. Finally, pretrial services implemented telephone (instead of in-person) check-ins for some individuals on monitored pretrial release.^{viii}

The CJC's decarceral reform effort also included an investigation into whether the County should incorporate a risk assessment instrument into its predisposition process. In this context, a risk assessment instrument was a scoring system that assigned numbers, or "weights," to certain events (e.g., a defendant's prior convictions) or characteristics (e.g., a defendant's age) to compile a score. The score purportedly classified the risk that an arrested individual if released predisposition would do something undesired, such as be rearrested (particularly for a violent offense) or fail to appear at a subsequent court date. Decision makers could review an individual's score when deciding whether to release them before disposition and, if so, the type of conditions to impose.^{ix} Conditions included how much and what type of bail and what kind (if any) of monitoring. If implemented, a risk assessment instrument's scores would become one among several factors Dane County decision makers could review. Other factors included a law enforcement affidavit/police report, information about the individual's ties to the community, and specifics of the individual's criminal history. Key decision makers who might use a risk assessment instrument included the Commissioner, who made the initial decision regarding predisposition release at the in-custody first appearance hearing; the judge, who oversaw the case from first appearance hearing to disposition, and who had the power to change predisposition release conditions (including those regarding bail); and Pretrial Services, which had a role in assigning the frequency and type of contact/check-ins for released individuals ordered to the County's monitoring program.

Around 2015, the CJC recommended that the County incorporate a risk assessment instrument into its predisposition process, but to do so cautiously. At this time, risk assessment instruments were controversial, especially in criminal justice. Detractors argued that they worsened racial disparities, caused more people to be incarcerated, or lessened individualized consideration of each defendant. Proponents, in response, contended that they reduced racial disparities or decreased incarceration. Proponents also argued that risk assessment instruments improved

judicial decisions, thereby lowering rates of undesired behaviors such as failure to appear (FTA), rearrest for a new offense (NCA), or rearrest for a new violent offense (NVCA) during the predisposition period.

In the face of such controversy, and particularly cognizant of concerns regarding racial disparities, the County took an evidence-based approach. When in mid-2017, the Dane County courts began using a risk assessment instrument, the CJC also launched a gold-standard evaluation of the instrument's effects. The gold-standard evaluation took the form of an RCT, the type of study used to assess whether drugs and vaccines can be provided safely to the public. Specifically, the County did not immediately make the risk assessment instrument available to the Commissioner and other decision makers in all cases. Rather, in partnership with the A2J Lab, it made the instrument available in a randomly selected half of cases and proceeded without the instrument, i.e., just as it did before mid-2017, in the other half. Arrested individuals received notice of the study in the form of a short flyer. The County provided data on criminal justice outcomes for all cases to the A2J Lab for analysis.

The particular risk assessment the CJC chose was called the Public Safety Assessment-Decision Making Framework System, a product of research funded by AV, formerly the Laura and John Arnold Foundation.^x The PSA-DMF consisted of two parts. The first part, the PSA, took data on the arrested individual's criminal history, current charge, and age as inputs to produce two risk scores ranging from 1 to 6, with higher numbers signaling higher risk. These two 1-6 risk scores corresponded to risk of NCA and FTA. The PSA also produced a "flag," with a value of 1 signaling an elevated risk of NVCA and 0 meaning no such elevated risk.^{xi} The second part, the DMF, incorporated the PSA scores with community-specific determinations, policies, and values; state statutory requirements; and jurisdictional resources. The result was a recommendation regarding the conditions of release, including bail and supervision level. The PSA scoring system was the same in all jurisdictions, but the DMF recommendation system differed across jurisdictions. The decision on release and conditions rested always with the judicial official. The PSA did not produce a recommendation, and the DMF's was not binding. A redacted sample of the PSA-DMF printout from a Dane County case appears below.



DANE COUNTY CLERK OF COURTS
Public Safety Assessment – Report

215 S Hamilton St #1000
Madison, WI 53703
Phone: (608) 266-4311

Name: [REDACTED]

Spillman Name Number: [REDACTED]

DOB: [REDACTED]

Gender: Male

Arrest Date: 03/25/2017

PSA Completion Date: 03/27/2017

New Violent Criminal Activity Flag

No

New Criminal Activity Scale

1	2	3	4	5	6
---	---	---	---	---	---

Failure to Appear Scale

1	2	3	4	5	6
---	---	---	---	---	---

Charge(s):

961.41(1)(D)(1) MFC DELIVER HEROIN <3 GMS F 3

Risk Factors:

1. Age at Current Arrest
2. Current Violent Offense
 - a. Current Violent Offense & 20 Years Old or Younger
3. Pending Charge at the Time of the Offense
4. Prior Misdemeanor Conviction
5. Prior Felony Conviction
 - a. Prior Conviction
6. Prior Violent Conviction
7. Prior Failure to Appear Pretrial in Past 2 Years
8. Prior Failure to Appear Pretrial Older than 2 Years
9. Prior Sentence to Incarceration

Responses:

23 or Older
No
No
No
Yes
Yes
Yes
2
0
Yes
Yes

Recommendations:

Release Recommendation - Signature bond

Conditions - Report to and comply with pretrial supervision

Figure 1: Sample PSA-DMF A redacted PSA-DMF printout appears above. This arrested individual was charged with manufacturing, distributing, or delivering three grams or less of heroin. The information in the top rectangle shows (i) the absence of the PSA NVCA flag, indicating no elevated risk; (ii) a PSA NCA score of 4; and (iii) a PSA FTA score of 3. The middle rectangle shows the PSA inputs and their values. The bottom rectangle shows the DMF recommendation of a signature bond with a referral to Pretrial Services for supervision.

As mentioned above, the County launched the PSA-DMF in mid-2017. With generous AV financial support, the County and the A2J Lab simultaneously launched the RCT. From mid-2017 until December 31, 2019, the Commissioner, the prosecutor, and the defense attorney at each in-custody first appearance hearing had access to the PSA-DMF printout for a randomly selected half of cases, and no such access for the other half. The CJC and the A2J Lab followed study cases until December 31, 2021. The CJC provided final data to the A2J Lab for analysis over the ensuing months. This report summarizes the A2J Lab's findings.

III. Changes in Decisions But No Broader Effect (Good or Bad)

The gist of the RCT study was a comparison of what happened in the cases in which the PSA-DMF was available to the presiding Commissioner (and the attorneys) versus what happened in the cases in which the PSA-DMF was unavailable. Because the assignment of a case to one group or the other was random, the two groups were as similar as possible in all ways except whether the PSA-DMF was available. Any statistically significant differences in results between the two groups, meaning differences unlikely to be due to chance, were thus attributable to the availability of the PSA-DMF. The A2J Lab examined the data to assess differences, if any, in six areas: Commissioner decisions; FTA; NCA; NVCA; number of days of predisposition incarceration; and measures of racial and gender fairness.

A word regarding terminology: when a Commissioner at an in-custody first appearance hearing assigned a defendant a “signature bond,” then for the purposes of that case,^{xii} the arrested individual could be released predisposition without the requirement of depositing money (“bail”) with the County. A “cash bail” decision meant that, to achieve release, the individual had to deposit the amount specified by the County, with the amount subject to seizure if the individual violated the conditions of release. For the other study sites, both of which featured bail bonds industries, the corresponding terms were “release own recognizance (ROR)” and “money bond.”

In Dane, during the study period, just over 70% of in-custody first appearance hearings resulted in signature bond decisions. About half of the remaining 30% received cash bail of \$500 or less (“low bail”). The other remaining half received cash bail of more than \$500 (“high bail”).

A. Commissioner Decisions: Moderate Effect

The availability of the PSA-DMF caused modest changes in Commissioner decisions, and the changes were in the expected direction. In other words, when the PSA’s NCA, NVCA, or FTA scores indicated higher risk, the availability of the PSA-DMF printout caused Commissioners to assign cash bail (sometimes low, sometimes high) more often than in cases when the printout was not available. In contrast, when the NCA, NVCA, or FTA scores indicated lower risk, for the most part, the availability of the PSA-DMF printout caused no or small changes, with some increased assignment of signature bonds or of low bail. Most differences were around 5-10 percentage points and were not statistically significant. A few were 11-14 percentage points, and there was one 24-percentage-point difference (corresponding to individuals receiving an FTA score of 5); these were all statistically significant.

The following tables illustrate. Table 1 shows the fraction of cases in which the Commissioner assigned a signature bond, low cash bail (less than or equal to \$500), and high cash bail (greater than \$500) for individuals receiving an NCA score of 5. Here, there is a change of 13 percentage points. Table 2 shows the same decision fractions for individuals receiving an FTA score of 1. Here, there is a change of four percentage points.

Table 1: Commissioner Decisions (Fractions) by Treatment Condition, NCA Score of 5

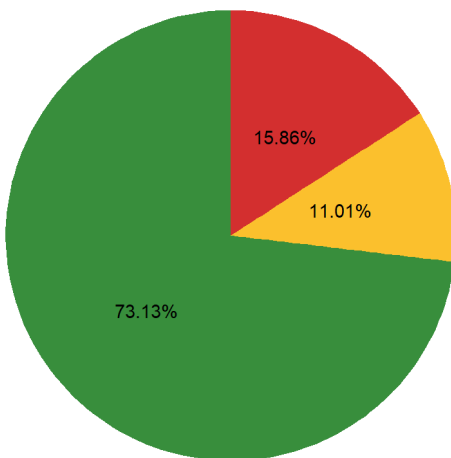
NCA = 5, Fraction of Cases	Signature Bond	Low Cash Bail	High Cash Bail
PSA-DMF Available	.45	.30	.25
PSA-DMF Not Available	.58	.19	.23
Difference	-.13	.11	.02
Statistically Significant?	Yes	Yes	No

Table 2: Commissioner Decisions (Fractions) by Treatment Condition, FTA Score of 1

FTA = 1, Fraction of Cases	Signature Bond	Low Cash Bail	High Cash Bail
PSA-DMF Available	.91	.03	.06
PSA-DMF Not Available	.88	.02	.10
Difference	.03	.01	-.04
Statistically Significant?	No	No	No

Overall, the PSA-DMF group experienced a three-percentage-point reduction in the frequency of signature bonds, a statistically significant difference. The A2J Lab does not consider this reduction of great importance for two reasons. First, three percentage points does not seem substantively large. Second, as described below, the lower signature bond rate in the PSA-DMF group did not translate into a statistically significant difference in predisposition incarceration days. The PSA-DMF group actually experienced a lower average number of predisposition incarceration days, although again, this effect was not statistically significant.

Distribution of Judicial Decisions
When PSA-DMF Not Present



Distribution of Judicial Decisions
When PSA-DMF Present

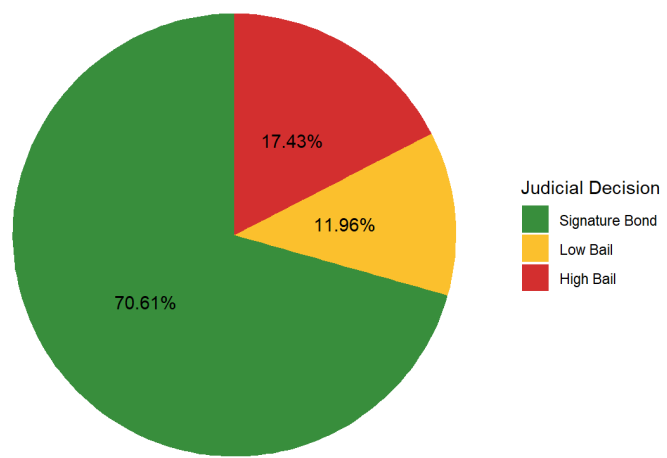


Figure 2: Commissioner Decisions by Treatment Condition *These pie charts show rates for signature, low cash bail, and high cash bail by treatment condition. Overall, the PSA-DMF group experienced a statistically significant three percentage point drop in Signature Bonds.*

To place the above figures in context: as noted above, the A2J Lab has completed a similar RCT evaluation of the PSA-DMF in Polk County. There, the availability of the PSA-DMF caused differences as high as 35-40 percentage points at certain PSA score levels, with most changes statistically significant. Such was not the case in Dane; the availability of the PSA-DMF caused only moderate differences.

B. NCA, NVCA, and FTA: No Differences

The purpose of the PSA-DMF was not just to change Commissioner decisions. Rather, the hope was that the availability of the PSA-DMF printout would change Commissioner decisions so as

to reduce some or all of NCA, NVCA, FTA, the number of days of predisposition incarceration, or measures of racial or gender disparity. No such reductions emerged. The moderate changes in Commissioner decisions did not translate into improved or worsened criminal justice outcomes. Instead, there were no statistically significant effects. Figures 3-5 illustrate.

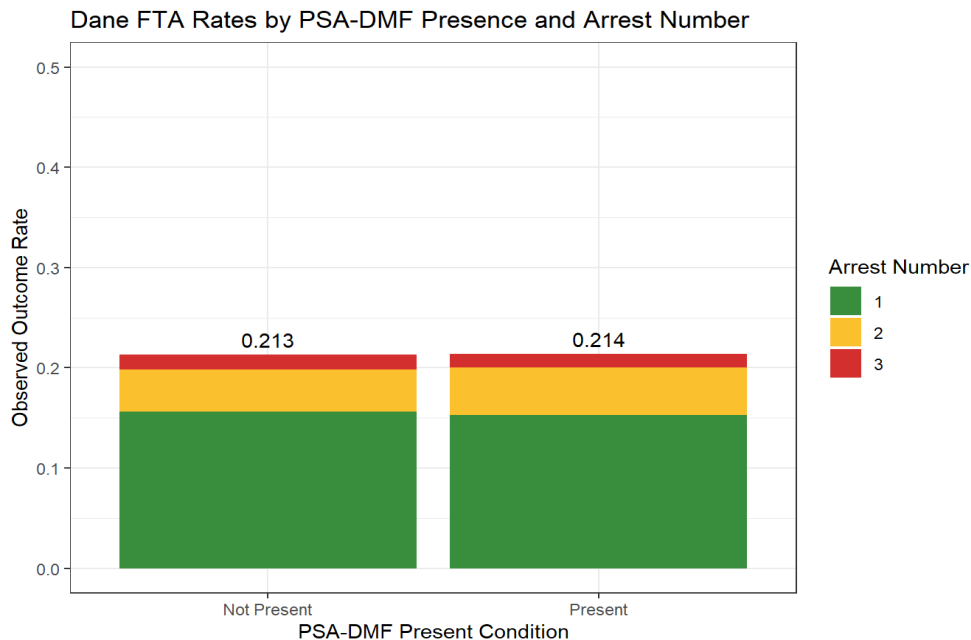
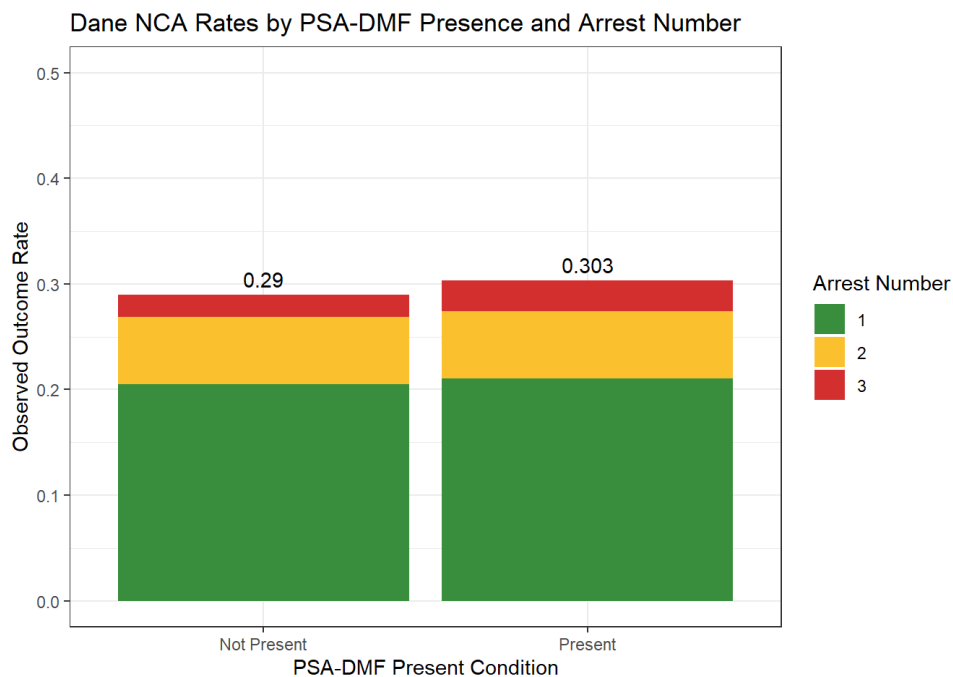


Figure 3: FTA Comparison

This figure shows the fraction of not-PSA-DMF cases (left column) and PSA-DMF cases (right column) experiencing at least one FTA in an arrestee's first arrest (green), second arrest (yellow), and third arrest (red). Totaled FTA rates (.213 and .214) were nearly identical. The PSA-DMF had no effect.

Figure 4: NCA Comparison

This figure shows the fraction of not-PSA-DMF cases (left column) and PSA-DMF cases (right column) experiencing at least one NCA in an arrestee's first arrest (green), second arrest (yellow), and third arrest (red). Totaled NCA rates (.290 and .303) were nearly identical. The PSA-DMF had no effect.



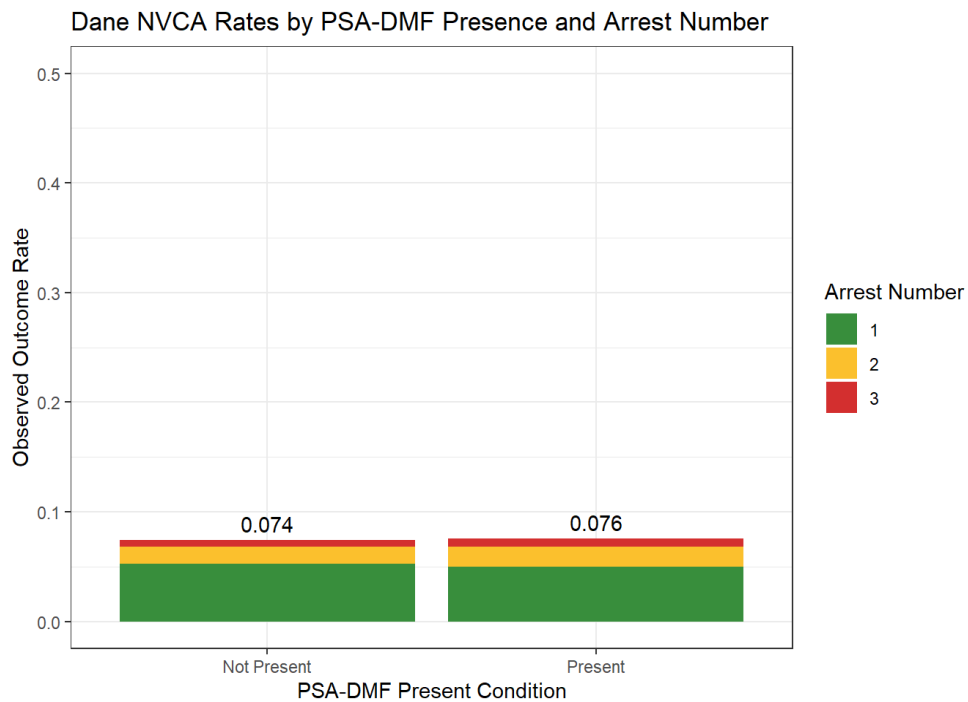


Figure 5: NVCA Comparison

This figure shows the fraction of not-PSA-DMF cases (left column) and PSA-DMF cases (right column) experiencing at least one NVCA in an arrestee's first arrest (green), second arrest (yellow), and third arrest (red). Totalled NVCA rates (.074 and .076) were nearly identical. The PSA-DMF had no effect.

As noted above, the results were similar across all three RCTs. Figure 6 depicts the three above graphs for the Dane,^{xiii} Polk, and Utah sites. There was no pattern of relevant differences in criminal justice outcomes. Although the difference in FTA rates in Polk, and in NVCA rates in Utah, met traditional levels of statistical significance, the absence of a pattern across the three sites as well as the modest effect sizes overall suggests that these differences have little substantive importance.

Outcome Rates by PSA-DMF Presence and Arrest Number

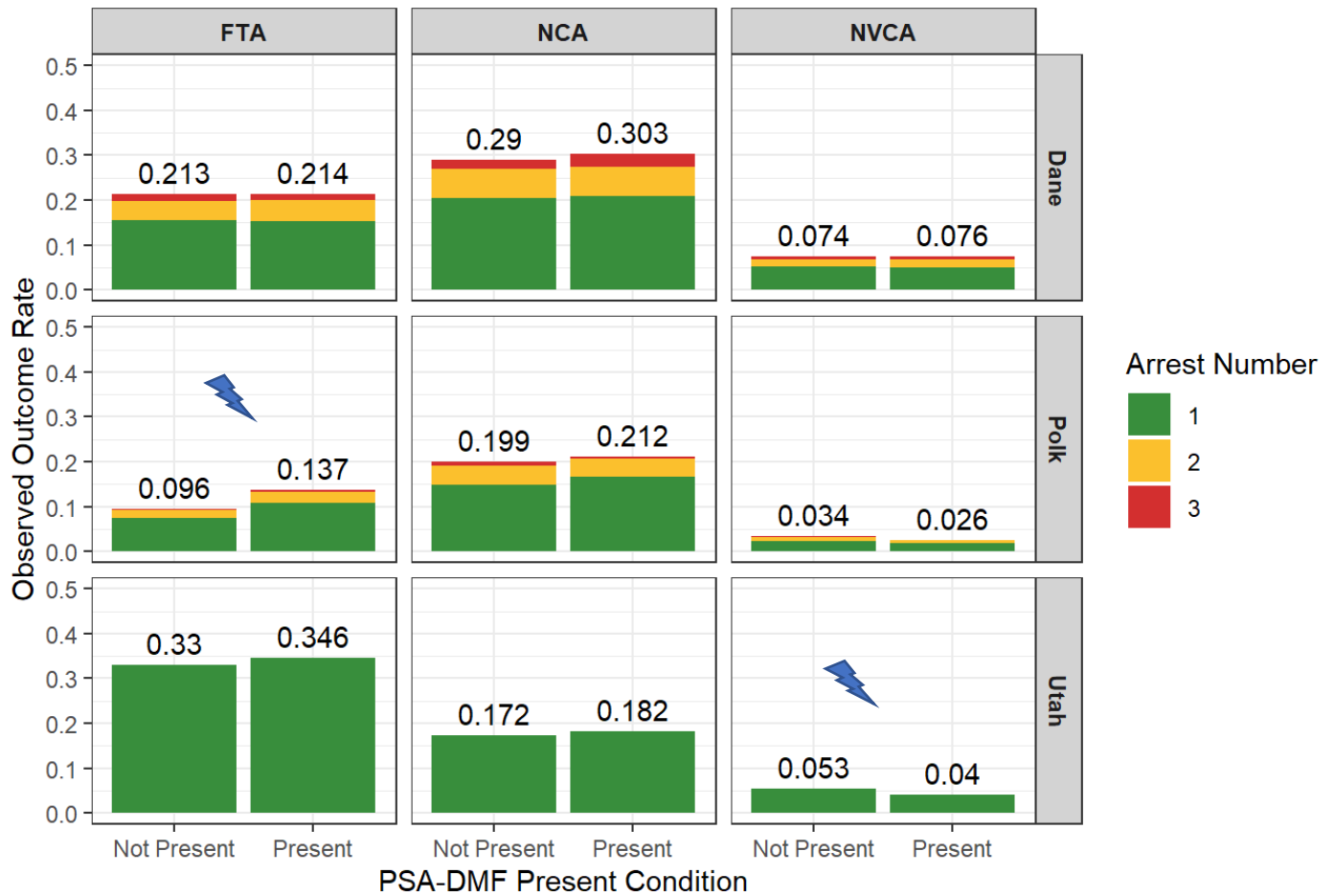


Figure 6: All Sites, All Comparisons This figure reproduces Figures 3-5 for all three study sites. The blue lightning bolts identify the two comparisons that produced differences meeting traditional thresholds of statistical significance. There was no pattern of statistically significant differences in FTA, NCA, or NVCA across the three sites, and no statistical significance in Dane.

C. Number of Days of Predisposition Incarceration: No Difference

Results were similar for the number of days of predisposition incarceration. Figure 7 graphs the distribution of predisposition days incarcerated for the PSA-DMF group versus the no-PSA-DMF group. The two distributions were similar, with the overwhelming majority of arrested individuals experiencing incarceration time periods on the short end of the distribution. There was no statistically significant difference between the PSA-DMF group and the no-PSA-DMF group.

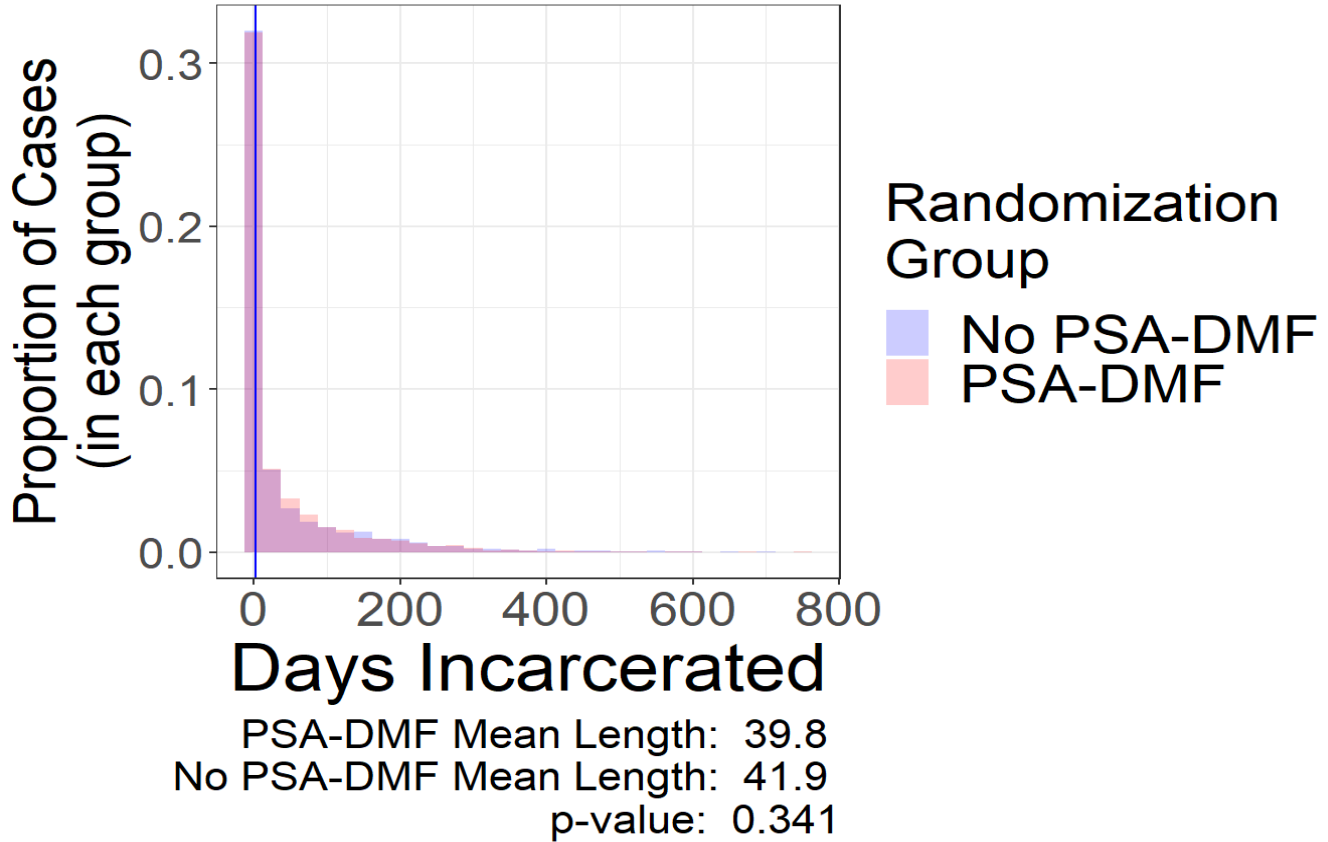


Figure 7: Comparison, Number of Days of Predisposition Incarceration

This figure shows the distribution of number of days of predisposition incarceration for the not-PSA-DMF group (in light blue) versus the PSA-DMF group (in light red). Areas of purple indicate overlap in the distributions. Almost the entire graph is purple, suggesting no difference between the two groups. Average incarceration durations (39.8 days for PSA-DMF, 41.9 for not-PSA-DMF) showed no statistically significant difference.

D. Measures of Racial and Gender Fairness: No Differences

Results were similar for measures of racial (and gender) fairness. Depicting results for racial and gender fairness comparisons is complicated because there are dozens of fairness measures. Some are mathematically incompatible with others. Most are designed for risk assessment instruments that have only two levels, corresponding in a loose sense to “high” and “low” risk. While the PSA’s NVCA flag has two levels, the NCA and FTA scales have values between 1 and 6. The A2J Lab addressed these issues by calculating racial and gender fairness results for several such metrics and for all possible ways of transforming a 1-6 measure into two levels. The A2J Lab then compared the metrics for the PSA-DMF group versus the no-PSA-DMF group, looking for patterns of statistically significant differences. Sample results appear in Figure 6, which graphs statistics known as “p-values” for comparisons of the PSA-DMF and no-PSA-DMF groups. Generally speaking, p-values *below* 0.05 indicate statistical significance. There were few statistically significant differences, and no patterns emerged. There was thus no evidence that PSA-DMF availability improved or worsened racial or gender disparities.^{xiv}

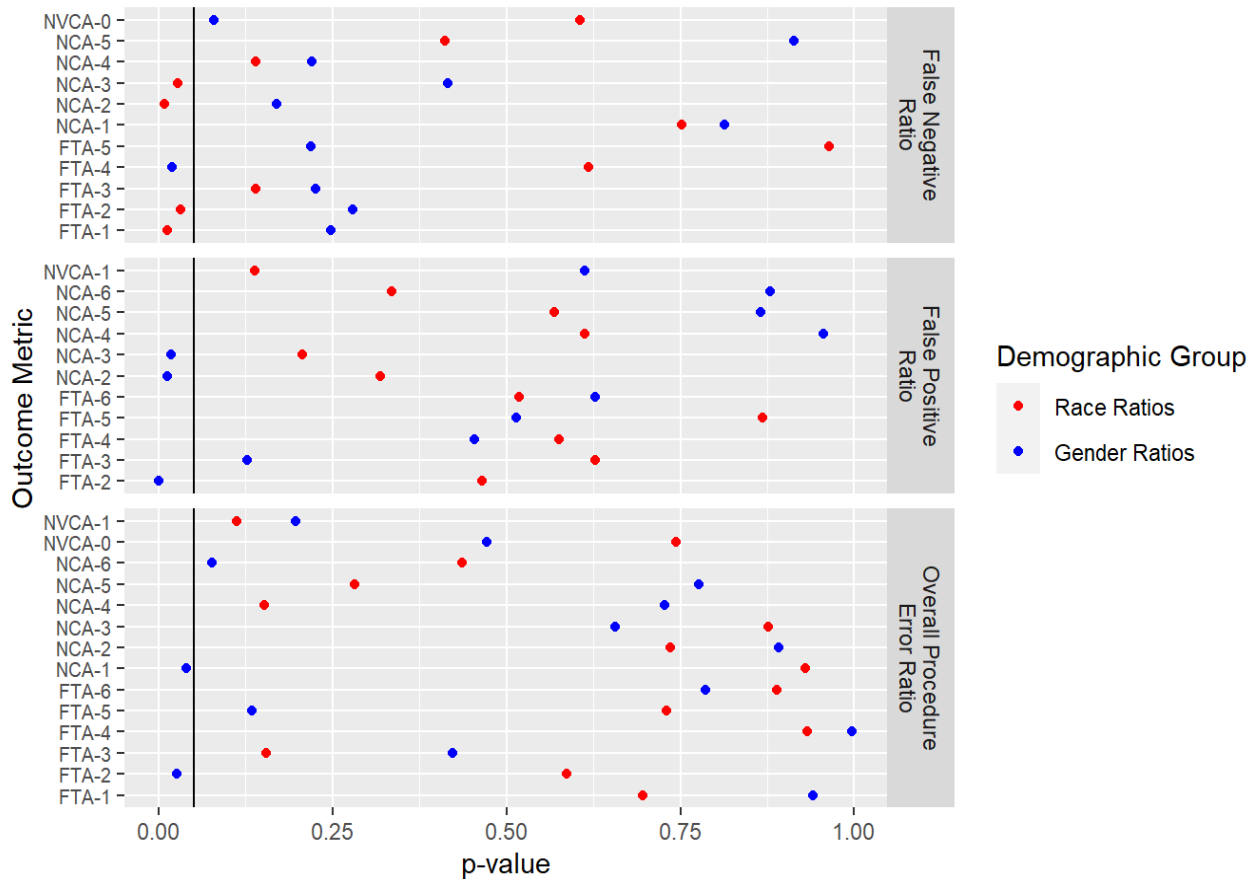


Figure 8: Racial and Gender Fairness Comparisons *This figure depicts p-values for comparisons of racial (red) and gender (blue) metrics of fairness across PSA-DMF versus no-PSA-DMF groups. Patterns of values less than 0.05 would indicate an improvement or worsening of racial or gender disparities. No such patterns were evident.*

IV. Explanations for the Absence of Effects

As suggested several times above, the fundamental result of this study was that, while the availability of the PSA-DMF caused moderate changes in Commissioner decisions, there were none of the (hoped-for) reductions in NCA, NVCA, FTA, number of days of predisposition incarceration, or measures of racial disparity. Instead, for the most important criminal justice outcomes, there were no relevant statistically significant differences in the PSA-DMF versus no-PSA-DMF groups. This section discusses why such might have occurred, i.e., why no improvements in criminal justice outcomes emerged.

The A2J Lab considers three explanations worth discussing, the first unlikely, the second and third likely. First, the availability of the PSA-DMF may not have caused large enough changes in Commissioner decisions. Although this explanation is possible, it is inconsistent with results from the Polk County study, and the A2J Lab does not consider it likely. Second, arrested individuals did not respond to changes in Commissioner decisions, i.e., assigning a signature bond versus low bail versus high bail. In other words, for all but a handful of arrestees, cash bail lacked the power to prevent misbehavior. Third, the PSA did not have classify arrestee risk well.

The A2J Lab considers the second (especially) and the third explanations likely. Both are consistent with the results of the Polk and Utah sites. And since the A2J Lab's 2024 report, the statistical case for these explanations has strengthened.

A. Unlikely: Insufficient Changes in Commissioner Decisions

As noted above, the availability of the PSA-DMF printout caused moderate changes in Commissioner decision making. Most changes in the assignment of a signature bond, low bail, or high bail were not statistically significant, and those that were tended to be approximately 5-15 percentage points in size. The PSA-DMF could not change the behavior of an arrested individual in terms of engaging in FTA, NVCA, or any other outcome directly (the way, say, an antibiotic changes the progression of an infectious disease). Rather, the PSA-DMF could work only by changing Commissioner decisions. If the PSA-DMF did not change decisions enough, it could not change criminal justice outcomes.

While this explanation is possible, the A2J Lab's current best conclusion is that it is not the source of the observed lack of effect on FTA, NVCA, NCA, number of days of predisposition incarceration, and measures of racial or gender disparity. As noted above, the A2J Lab observed large changes in judicial officer decisions (sometimes approaching 40 percentage points) in Polk County, and modest differences in the Utah site. Table 3 depicts Polk magistrate decisions at in-custody first appearance for cases in which the FTA score equaled 1. Here, "OR Release" corresponds to a Dane Signature Bond. The shifts in decisions were statistically significant and substantively large.

Table 3: Polk Magistrate Decisions (Fractions) by Treatment Condition, NCA Score of 1

NCA = 1, Fraction of Cases	OR Release	Low Bond	High Bond	Remand
PSA-DMF Available	.65	.22	.12	.01
PSA-DMF Not Available	.29	.45	.24	.01
Difference	.35	-.23	-.12	.00
Statistically Significant?	Yes	Yes	Yes	No

Across all three sites, including Polk, there was no consistent pattern of changes in criminal justice outcomes. Because there were no relevant changes in these outcomes regardless of the size of the alterations in judicial officer decisions, the A2J Lab does not believe that the modest size of the changes in Commissioner decisions in Dane is a sufficient explanation for the lack of an effect in criminal justice outcomes.

B. Likely: Ineffectiveness of Bail, Little Classification Power

The best way of understanding both likely explanations for the lack of a PSA-DMF effect depends on concepts the A2J Lab and Professor Imai's team adapted over the past several years from statistical literature. Those techniques allowed the research team to estimate whether a decision to impose bail prevents negative criminal justice outcomes during the predisposition period. The techniques worked by estimating the fraction of individuals falling into each of the following three categories.

- “**Safe**” individuals would not commit FTA or N(V)CA regardless of Commissioner decision. In other words, “Safe” individuals would not misbehave if the Commissioner assigned a signature bond or cash bail.
- “**Preventable**” individuals would commit FTA or N(V)CA if the Commissioner assigned a signature bond but would avoid doing so if the Commissioner assigned cash bail. In other words, for “Preventable” individuals, a decision to impose cash bail would prevent misbehavior, either by incapacitating the arrested individual (because the individual could not post the assigned bail amount and thus remained in jail) or by deterring the arrestee (who achieved release by posting the assigned bail but avoided misbehavior for fear of forfeiting the posted amount).
- “**Hopeless**”^{xv} individuals would commit FTA or N(V)CA regardless of Commissioner decision. In other words, “Hopeless” individuals would misbehave if the Commissioner assigned a signature bond or cash bail.^{xvi}

It is not possible to identify with certainty at the time of a bail decision whether an arrested individual falls into the Safe, Preventable, or Hopeless categories.^{xvii} But the advances in statistical techniques from the A2J Lab and Imai teams allowed several conclusions, three relevant here. *First*, it was possible, after an RCT, to estimate the fraction of Safe, Preventable, and Hopeless arrestees in a jurisdiction. *Second*, for a quantitative guidance tool to inform judicial decision making to have any work to do, there must be a reasonably high fraction of Preventable arrestees. If the fraction of Preventable arrestees is low, then the decision to impose bail will rarely prevent FTA or N(V)CA. *Third*, although it is not possible before the fact (meaning at a bail decision) to identify with certainty whether an arrestee is Safe, Preventable, or Hopeless, it is at least theoretically possible for a quantitative guidance tool to provide some information about the *probability* or *likelihood* that an arrestee is in one category or the other.

Such should be the goal of a guidance tool. To repeat, the goal of an instrument designed to guide the signature versus cash decision should be to provide information about the probability or likelihood that an arrestee is Safe, Preventable, or Hopeless. In particular, the guidance instrument should inform as to the likelihood that an arrestee is Preventable. That is because, in a broad way, the decision to impose cash bail makes sense only for Preventable arrestees. If a quantitative guidance tool does a poor job differentiating (on average) Safe from Preventable from Hopeless, then the tool is of little, if any, use.

These concepts inform the remainder of this report. Following the first point in the previous paragraph, the A2J Lab ran an RCT in Dane County, along with counterpart RCTs in Polk

County and in Utah. Following the previous paragraphs' second point, (very) few arrestees were Preventable in all three sites, including Dane. Following the previous paragraph's third point, in all three sites (including Dane), the PSA provided (very) little useful information on the probability that an arrestee fell into the Safe, the Preventable, or the Hopeless category.

1. Imposing Bail Is Effective for Only a Small Fraction of Arrestees

The A2J Lab's estimates that only a small fraction of arrestees were Preventable. The figures are the same in Dane, Polk, and Utah. For reasons explained below, the A2J Lab's confidence in these results is highest in Utah, but the results are similar across the three sites, increasing the A2J Lab's confidence that the nearly unassailable results in Utah also apply to Dane (and Polk).

To explain, the key assumption underlying the conclusion that so few arrestees were Preventable was that the research team observed enough of the information that the judicial decision maker had available at the time of the bail decision. In Dane and Polk, there was information to which the research team lacked access, namely, the probable cause affidavits supporting each arrest and the colloquy between the judicial officer and others present at each in-custody first appearance hearing. The assumption, then, was that the research team possessed "enough" of the key information about each arrestee despite these lacks. In Utah, there was no lack. There, the A2J Lab, after months of work, obtained the probable cause affidavits for all study cases. And in the Utah system, judicial officers make bail decisions in the absence of hearings, so there were no colloquies. Thus, the A2J Lab had access to all information available to judicial officers in Utah but not in Dane or Polk. The confidence for Dane and Polk thus rested on the untestable assumption that the research team had access to enough information, as well as on the similarity of the results across the three sites.^{xviii}

Figures 9-11 depict the estimated fraction of Safe, Preventable, and Hopeless cases under the assumptions discussed above in Dane County for FTA, NCA, and NVCA.

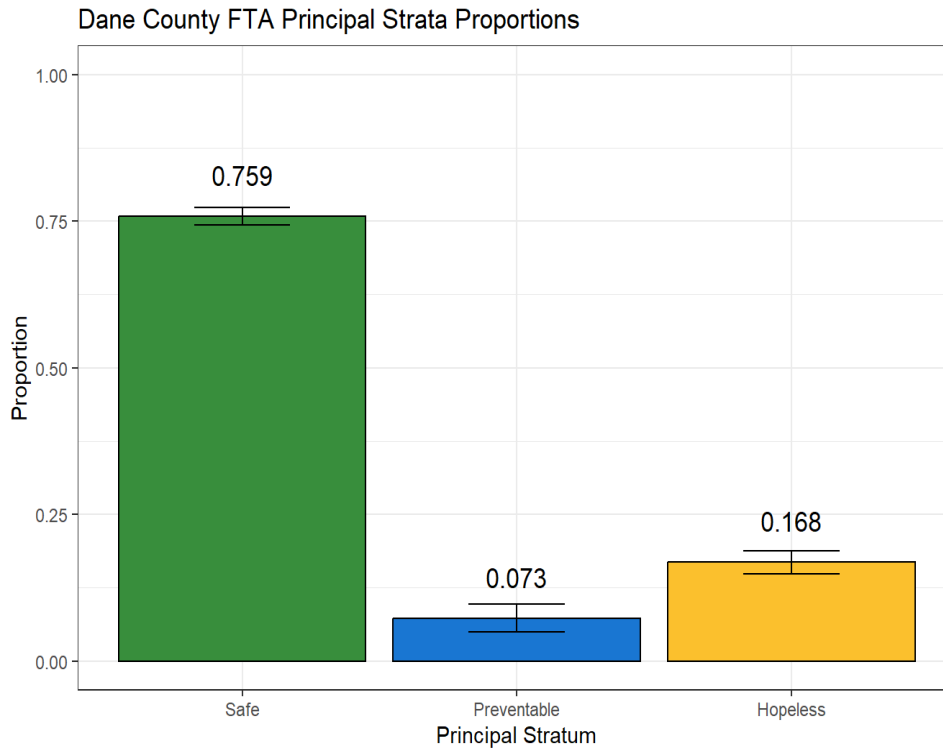
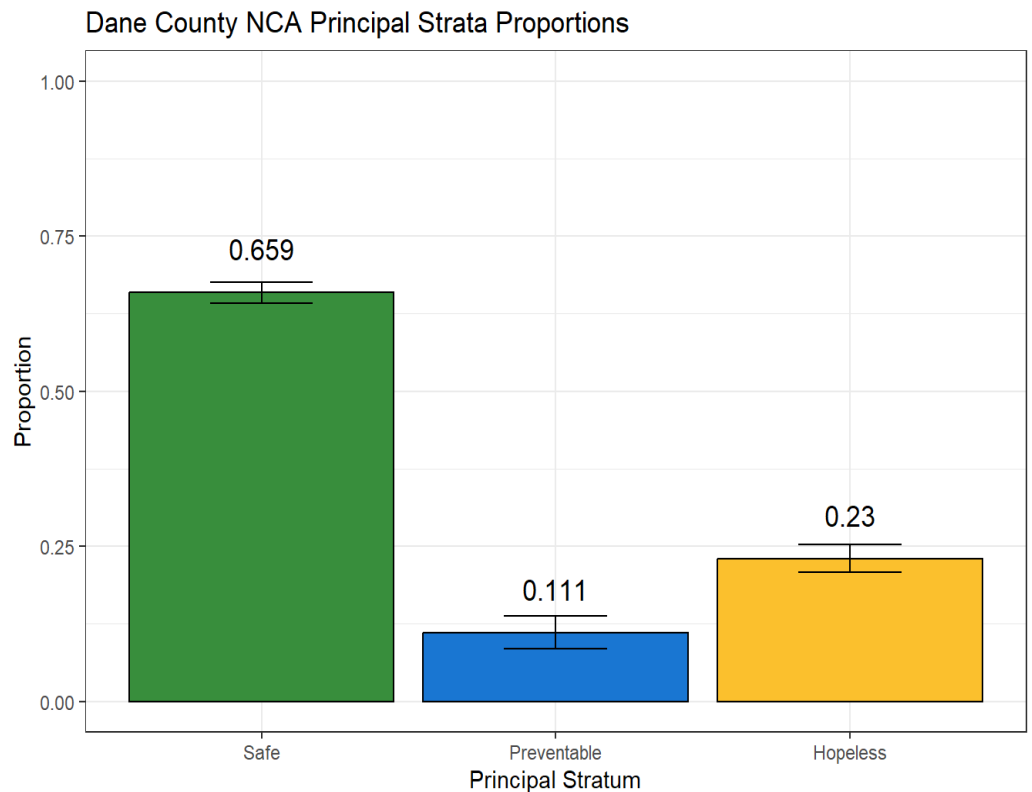


Figure 9: FTA Proportions This figure depicts the fraction of Safe (green), Preventable (blue), and Hopeless (yellow) cases in Dane for FTA. Imposing bail can affect only Preventable cases, but the estimated fraction of such cases is only about 7%. Moreover, the results from Polk and Utah suggest that this 7% figure is probably misleadingly high.

Figure 10: NCA Proportions This figure depicts the fraction of Safe (green), Preventable (blue), and Hopeless (yellow) cases in Dane for NCA. Imposing bail can affect only Preventable cases, but the estimated fraction of such cases is only about 11%. Moreover, the results from Polk and Utah suggest that this 11% figure is probably misleadingly high.



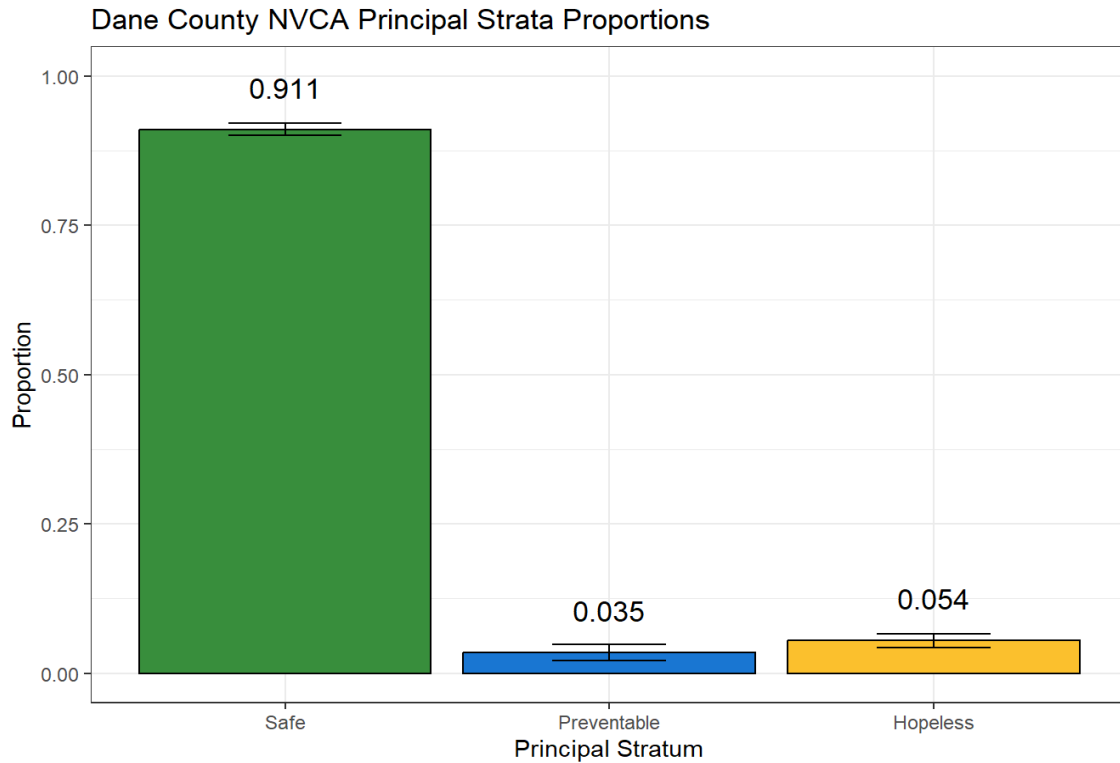


Figure 11: NVCA Proportions *This figure depicts the fraction of Safe (green), Preventable (blue), and Hopeless (yellow) cases in Dane for NVCA. Imposing bail can affect only Preventable cases, but the estimated fraction of such cases is less than 4%. Moreover, the results from Polk and Utah suggest that this 4% figure is probably misleadingly high.*

As noted above, the results were similar across all three RCTs. Figure 12 depicts the three above graphs for the Dane, Polk, and Utah sites. In all three sites, and for all three outcomes, the fraction of Preventable cases was at or below twelve percent, usually much lower. For NVCA, the fraction was extremely low in all three locations.

Principal Strata Sizes

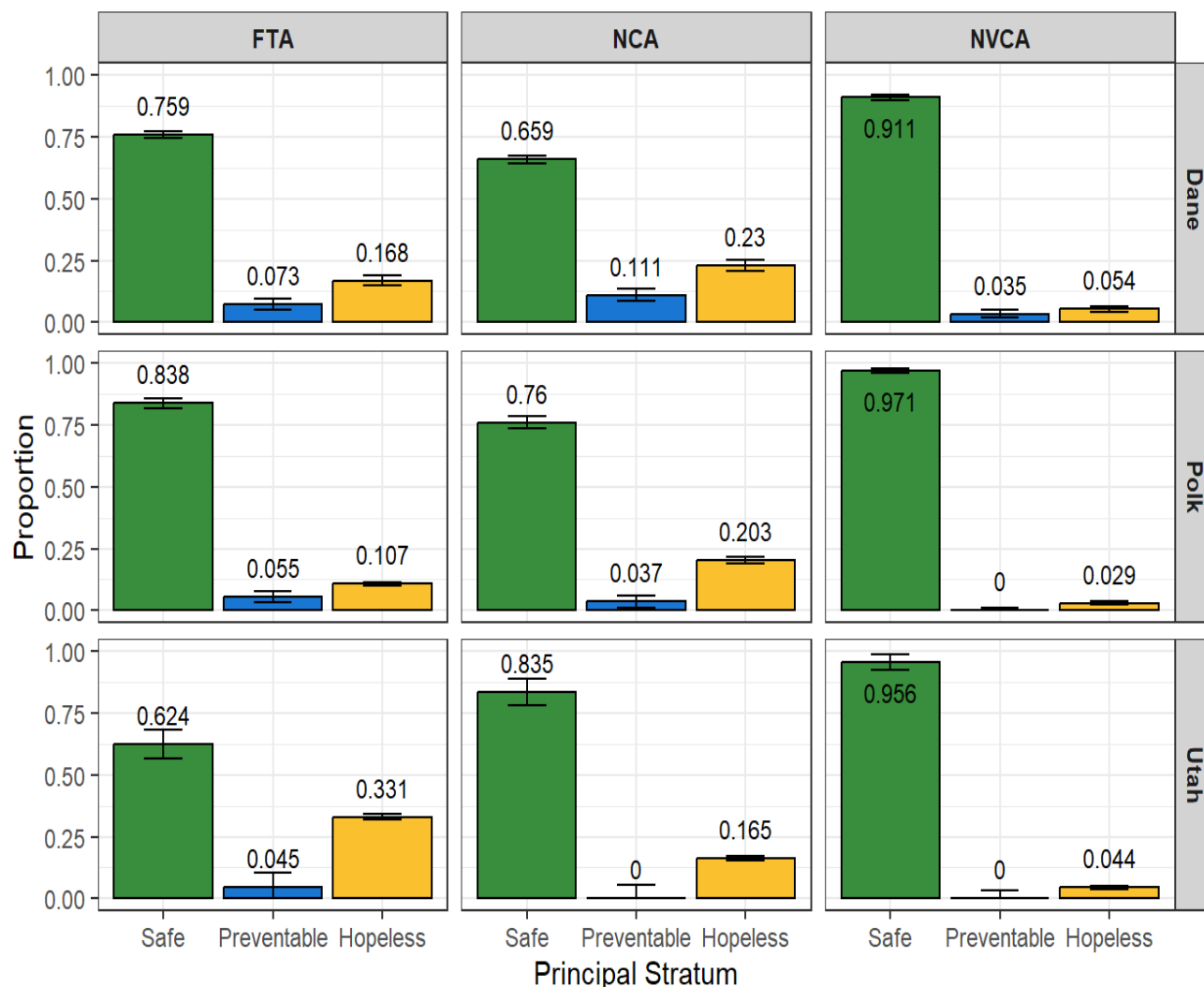


Figure 12: All Sites, Proportions This figure reproduces Figures 9-11 for all three study sites. Safe (green) arrestees will not misbehave regardless of Commissioner decision. Preventable (blue) arrestees will misbehave with signature but not with cash. Hopeless (yellow) arrestees will misbehave regardless of Commissioner decision. Imposing bail can affect only Preventable cases. The estimated fraction of Preventable cases for each of Dane, Polk, and Utah, for each of FTA, NCA, and NVCA, is at or below twelve percent, usually much lower. For NVCA, the fractions are extremely low.

2. Insufficient Classification Power

The availability of the PSA-DMF could change criminal justice outcomes only if its classifications were strong enough to provide the Commissioner with powerful information. For example, if an individual receiving a PSA FTA score of 4 was actually little different in terms of risk from an individual with a PSA FTA score of 6, the FTA scale likely did not provide the Commissioner with enough information to make decisions that would reduce FTA.

The A2J Lab's statistical tests suggest that, in fact, the PSA-DMF's classifications provided the Commissioner with little, perhaps even no, useful information. Figure 13 illustrates. This figure graphs the FTA rate, for each value of the PSA's 1-6 FTA scale, for cases in which the arrested individual spent at least one day out of jail during the predisposition periods. There was only a 22-percentage-point difference between the lowest risk category (1) and the highest risk category (6), meaning that on average each step up (1 to 2, 2 to 3, etc.) corresponded to about a 4-percentage-point increase in FTA rate. Jumps in score from, say, 1 to 3, or from 4 to 6, did not correspond to discernible differences in risk. More than half of the cases receiving a score of 6, the highest risk classification, did not include an instance of FTA.

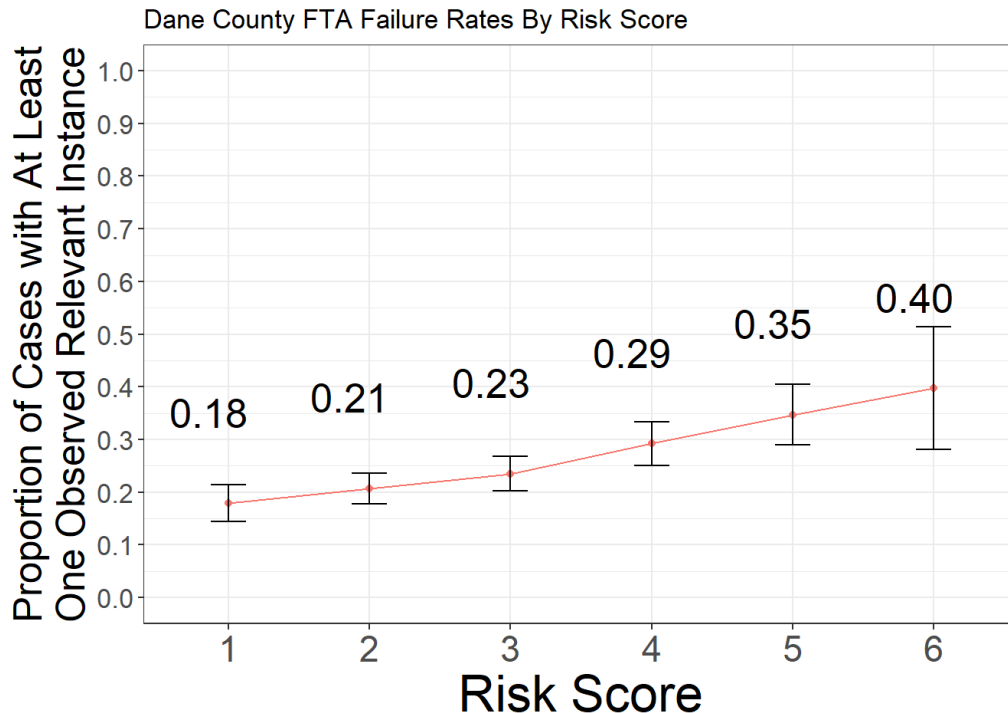


Figure 13: FTA Rates by PSA FTA Score *This figure shows the FTA rates by PSA FTA score for cases in which the arrested individual spent at least one predisposition day out of jail. Actual FTA rates for the scores were surprisingly similar, suggesting that the FTA score did not provide strong differentiating information about FTA risk.*

The same was true for all three sites for all three criminal justice outcomes. Figure 14 depicts the FTA, NCA, and NVCA rates for the Dane, Polk, and Utah sites. Only for NVCA in Dane and for FTA in Utah were there easily noticeable (but nevertheless modest and non-uniform) increases in misbehavior rates corresponding to increases in scores.

Failure Rates by Risk Score

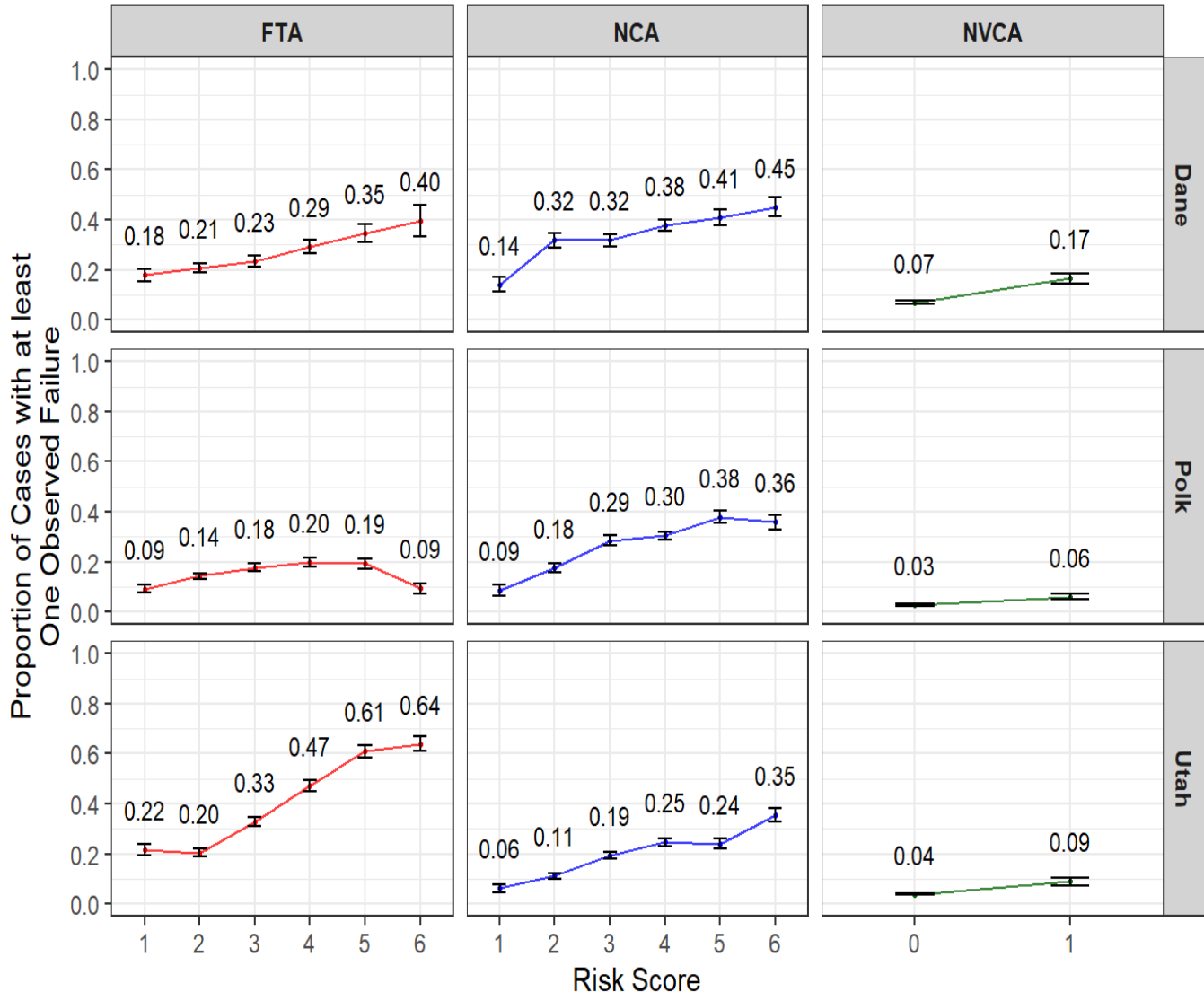
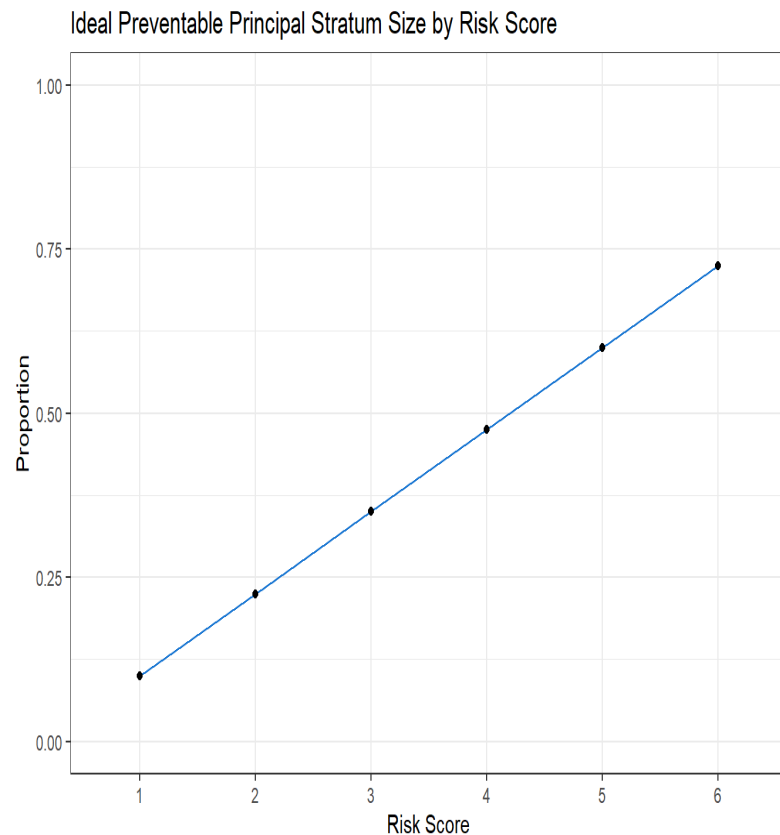


Figure 14: Failure Rates By Risk Score, All Sites *This figure reproduces Figure 13 for all three study sites and for all three criminal justice outcomes. Only for NVCA in Dane and for FTA in Utah were there easily noticeable (but nevertheless modest and non-uniform) increases in misbehavior rates corresponding to increases in scores.*

Graphs like those in Figures 13 and 14 are traditional in the business of validation of risk classification instruments. They are misleading for several reasons. For example, they do not account for the possibility that judicial officers could perhaps distort these graphs with well-placed imposition of bail (or, in jurisdictions that so allow, remand to jail), thereby incarcerating individuals with high misbehavior risk. More importantly, these graphs do not show whether the PSA provided useful information about the probability that an arrestee is Preventable. Per the discussion above, providing such information should be the purpose of any classification instrument.

More informative conclusions are now available. Recall, again, that the purpose of a classification instrument should be to provide information about the probability that each arrestee is Preventable. If the PSA were a useful classification instrument, then the fraction of preventable cases should (noticeably) increase as PSA scores increase. In other words, if the PSA provided useful information, the graph of fraction Preventable by PSA score would look something like Figure 15.

Figure 15: Desirable Classification Instrument *This figure shows what should happen for a hypothetical classification instrument that provides strong and useful information to a decision maker. Increases in instrument score corresponds to noticeable increases in the fraction of Preventable cases.*



In none of the three jurisdictions, for none of the three criminal justice outcomes, did the relevant graphs look anything like Figure 15. As Figure 16 depicts, the fraction of Preventable cases was indistinguishable at all levels of all PSA scores.

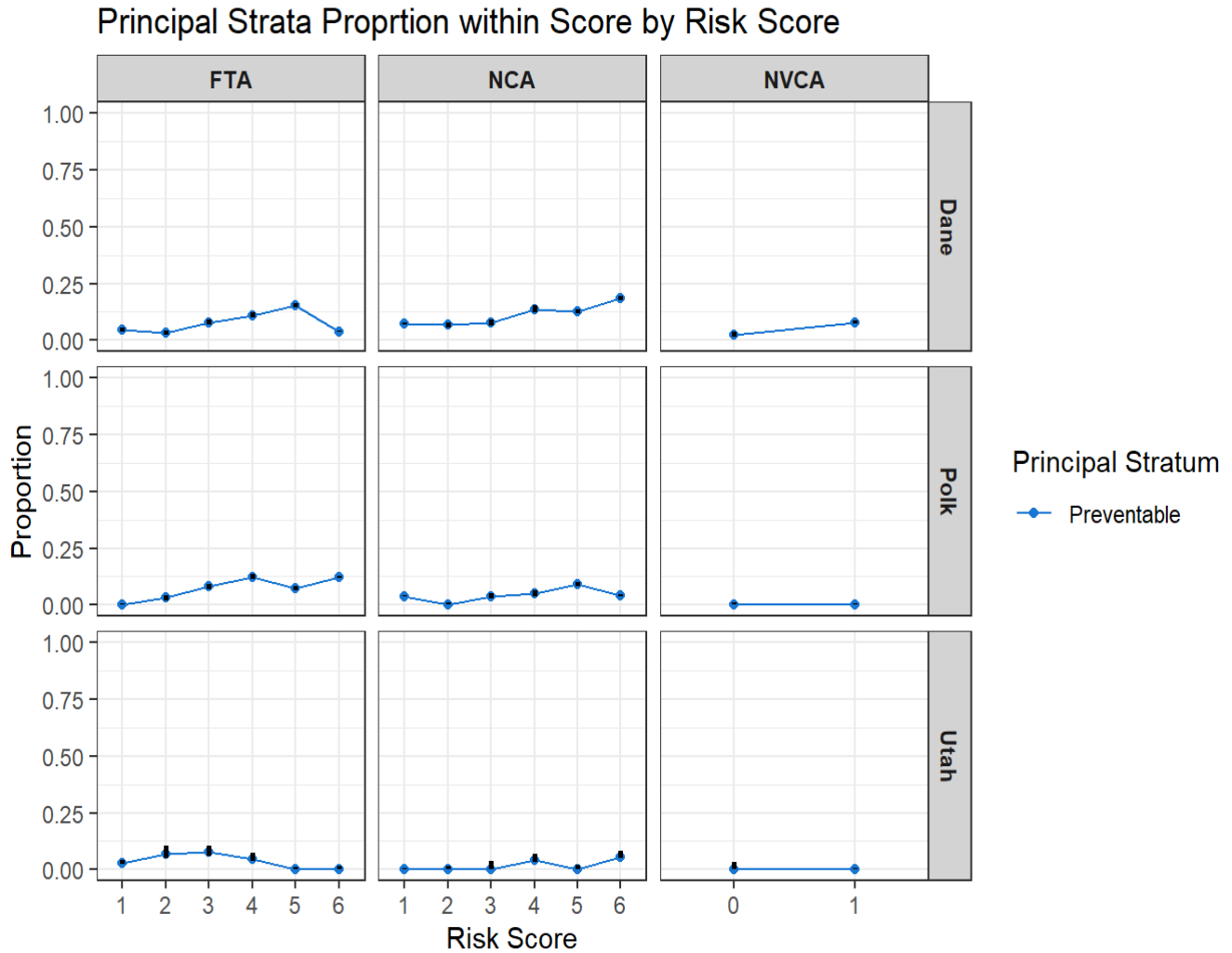


Figure 16: Classification Strength, All Sites, All Outcomes *This figure depicts the fraction of Preventable arrestees by PSA score for each of the three outcomes for each of Dane, Polk, and Utah. For no site and no outcome did the PSA identify Preventable arrestees well.*

IV. Conclusion

The CJC adopted the PSA-DMF as part of its overall effort to reduce the footprint of predisposition incarceration and to address perceived racial inequalities. The CJC did so cautiously, with an eye toward gold-standard assessment of the PSA-DMF's effects. Many, including the A2J Lab, hoped that the PSA-DMF would help to reduce some or all of FTA, NCA, NVCA, the number of predisposition incarceration days, and measures of racial disparity. Others feared that the PSA-DMF would worsen those same outcomes. The fact that neither those hopes nor those fears were realized shows the wisdom of the CJC's approach.

Many, perhaps most, efforts to make the world a better place fail. Failures are particularly common in criminal justice. But given the problems we face in the criminal justice system, complacency and inertia are not options. Trial and error is essential for improvement. The A2J Lab is grateful for the opportunity to participate in Dane County's improvement efforts.

ⁱ The four Utah Counties were Davis, Morgan, Utah, and Weber.

ⁱⁱ The CJC is a collaborative body “composed of executive level criminal justice leaders, including the Presiding Judge, Sheriff, District Attorney, Clerk of Courts, County Executive, and County Board Chair, as well as stakeholders from law enforcement, defense, municipal judiciary, and corrections.” It seeks to implement evidence-based policies to improve the County's criminal justice system.

<https://cjc.countyofdane.com/about-us> (last visited January 10, 2024).

ⁱⁱⁱ Pre-trial Services Subcommittee, Report to the Dane County Criminal Justice Council, available at <https://cjc.danecounty.gov/documents/pdf/2014-Report-to-the-Dane-County-Criminal-Justice-Council--Pre-trial-Services-Subcommittee--.pdf> (last visited July 13, 2026).

^{iv} Rule 206, Bail Hearings, <https://courts.countyofdane.com/prepare/CompleteRuleList#rule206> (last visited July 27, 2023).

^v Dane County CJC Research Team, Memo, “Comments/edits on the Interim Report “Randomized Control Trial Evaluation of the Implementation of the PSA-DMF System in Dane County, WI” (memorializing comments and suggested edits to the A2J Lab's Interim Report of Apr. 13, 2020) (on file with the Access to Justice Lab).

^{vi} Matthew Stubenberg, Memo, “Meeting with Clerk of Court and Register in Probate,” Memorializing Conversation on Dec. 12, 2019 (on file with the Access to Justice Lab).

^{vii} Email exchange, Manager of the Division of Policy and Practice Innovation Dane County Board of Supervisors & Matthew Stubenberg, Jul. 19, 2019 (on file with the Access to Justice Lab).

^{viii} Dane County CJC Research Team, Memo, *supra* note v.

^{ix} At this time, Wisconsin law required a demanding, and hardly-ever-used, process before an arrested individual could be remanded to jail without possibility of predisposition release. Thus, essentially all arrested individuals could achieve predisposition release if they could comply with ordered conditions.

^x [Maximize Opportunity. Minimize Injustice. | Arnold... | Arnold Ventures](#) (last visited December 28, 2023).

^{xi} A discussion of the PSA's inputs, initial integer weights, and processing of those weights into 1-6 FTA and NCA risk bands is available at <https://advancingpretrial.org/psa/factors/> (last visited Oct. 25, 2022).

^{xii} Not all individuals receiving signature bonds secured immediate release. An individual receiving a signature bond might, for example, be on pretrial release in another criminal matter, or be the subject of a probation or immigration hold. Such individuals might not be released immediately, or at all. Similarly, many individuals who received cash bail achieved predisposition release, perhaps by posting the assigned amount or by receiving a subsequent modification of the cash bail assignment.

^{xiii} The Dane results in Figures 3-6 are slightly different from those presented the A2J Lab's 2024 report and from those that appear 13-14. These differences are the result of slightly different ways of counting cases and of which subset of cases were included in the cases that generated each graph. Some of these differences result from the new statistical techniques the A2J Lab and Imai teams generated in the past two years. None of the differences are substantively significant.

^{xiv} The metrics deployed in Figure 8 are all based on a so-called “prediction matrix,” also called a “confusion matrix.” An explanation of this kind of matrix and accompanying metrics appears in the A2J Lab's interim Dane County report starting on page 67, see [Dane-County-PSA-DMF-RCT-Interim-Report.pdf \(a2jlab.org\)](#). In a nutshell, a confusion matrix compares predictions (e.g., “High Risk” versus “Low Risk”) to actual occurrences (e.g., “Bad Outcome” versus “Good Outcome”).

^{xv} A previous report explained these same concepts with two changes in language. First, the previous report used the term “Risky,” not “Hopeless,” to refer to individuals who would misbehave regardless of Commissioner decision. The A2J Lab has found that the term “Hopeless” is more communicative. Second, the previous report divided the “Preventable” category into “Easily Preventable” and

“Preventable” categories. Under the old terminology, either low or high cash bail, but not a signature bond, would stop an “Easily Preventable” arrestee from misbehaving; only high cash bail, but not a signature bond or low cash bail, would stop a “preventable” arrestee from misbehaving. The A2J Lab has combined these two categories into a single “Preventable” label for simplicity of presentation. Results using the old categories are the same as those presented here.

^{xvi} A fourth category is possible. “Backlash” arrestees are those who would commit FTA or N(V)CA if assigned cash bail but would not do so under a signature bond. Although it might at first appear counterintuitive that assigning cash bail would generate (as opposed to deter) criminal activity, the following example clarifies: a low-income parent who otherwise would have avoided misbehavior borrows money from an unsavory source to post cash bail, which induces them to steal money to pay back the source. For the analysis presented in this report, the A2J Lab has followed traditional bail analysis by assuming that Backlash arrestees do not exist. In other work, the A2J Lab and the Imai team have been investigating what happens if this assumption is relaxed. The A2J Lab is available to explain those results, which are more complex than those presented here, upon request.

^{xvii} After observing whether an arrested individual commits FTA or N(V)CA, it is possible to categorize definitively only some arrestees, and only if one assumes that no Backlash (see note xvi) individuals exist. For example, if the Commissioner imposes cash bail upon an arrestee and the arrestee nevertheless FTAs, then that arrestee (for FTA purposes) was either Backlash or Hopeless. If by assumption no Backlash individuals exist, then that arrestee must have been Hopeless.

^{xviii} The only caveat is with respect to the assumption, explained in note xvi, that no Backlash arrestees exist. The research team analyzed the data without this assumption. Oversimplifying somewhat, it turns out to be possible that the fraction of Preventable cases (i.e., those for which imposing bail prevents misbehavior) might have been higher than a small fraction, but only if the fraction of Backlash (i.e., those for which imposing bail generates misbehavior) cases was almost just as high.